Estimation of the functional Weibull-tail coefficient

Stéphane Girard

Inria Grenoble Rhône-Alpes & LJK, France
http://mistis.inrialpes.fr/people/girard/

June 2017

joint work with Laurent Gardes, Université de Strasbourg.

1. Weibull-tail distributions
2. Estimation of extreme conditional quantiles
3. Estimation of the functional Weibull-tail coefficient
4. Illustration on simulations

Notations

- Let $(X, Y) \in E \times \mathbb{R}$ be a random pair where E is an arbitrary space associated with a semi-metric (or pseudometric) d, see [3], Definition 3.2.
- The conditional survival function of Y given $X = x \in E$ is denoted by $\bar{F}(y|x) := \mathbb{P}(Y > y|X = x)$ and is supposed to be continuous and strictly decreasing with respect to y.
- The associated conditional cumulative hazard function is defined by $H(y|x) := -\log \bar{F}(y|x)$.
- The conditional quantile is given by $q(\alpha|x) := \bar{F}^{-1}(\alpha|x) = H^{-1}(\log(1/\alpha)|x)$, for all $\alpha \in (0, 1)$.

Conditional Weibull-tail distributions

(A.1) $H(.|x)$ is supposed to be regularly varying with index $1/\theta(x)$, *i.e.*

$$\lim_{y\to\infty} \frac{H(ty|x)}{H(y|x)} = t^{1/\theta(x)},$$

for all $t > 0$. In this situation, $\theta(.)$ is a positive function of the covariate $x \in E$ referred to as the functional Weibull tail-coefficient.

From [1], $H^{-1}(.|x)$ is regularly varying with index $\theta(x)$. Thus, there exists a slowly-varying function $\ell(.|x)$ such that

$$q(e^{-y}|x) = H^{-1}(y|x) = y^{\theta(x)}\ell(y|x).$$

Recall that the slowly-varying function $\ell(.|x)$ is such that

$$\lim_{y\to\infty} \frac{\ell(ty|x)}{\ell(y|x)} = 1, \tag{1}$$

for all $t > 0$.

Additional assumptions

(A.2) $\ell(.|x)$ is a normalised slowly-varying function.

In such a case, the Karamata representation (see [1]) of the slowly-varying function can be written as

$$\ell(y|x) = c(x) \exp\left\{ \int_1^y \frac{\varepsilon(u|x)}{u} du \right\},$$

where $c(x) > 0$ and $\varepsilon(u|x) \to 0$ as $u \to \infty$.

The function $\varepsilon(.|x)$ plays an important role in extreme-value theory since it drives the speed of convergence in (1) and more generally the bias of extreme-value estimators. Therefore, it may be of interest to specify how it converges to 0:

(A.3) $|\varepsilon(.|x)|$ is regularly varying with index $\rho(x) \leq 0$.

$\rho(x)$ is called the conditional second-order parameter.

Examples of (unconditional) Weibull-tail distributions

Distribution	θ	$\ell(y)$	$\varepsilon(y)$	ρ
Gaussian $\mathcal{N}(\mu, \sigma^2)$	$1/2$	$\sqrt{2}\sigma - \frac{\sigma}{2\sqrt{2}} \frac{\log y}{y} + O(1/y)$	$\frac{1}{4} \frac{\log y}{y}$	-1
Gamma $\Gamma(\alpha \neq 1, \lambda)$	1	$\frac{1}{\beta} + \frac{\alpha - 1}{\beta} \frac{\log y}{y} + O(1/y)$	$(1-\alpha)\frac{\log y}{y}$	-1
Weibull $\mathcal{W}(\alpha, \lambda)$	$1/\alpha$	λ	0	$-\infty$

Two goals

Starting from iid copies (X_i, Y_i), $i = 1, \dots, n$, of (X, Y),

- Estimate the extreme conditional quantiles defined as

$$\mathbb{P}(Y > q(\alpha_n, x) | X = x) = \alpha_n,$$

  when $\alpha_n \to 0$ as $n \to \infty$.

- Estimate the functional Weibull-tail coefficient $\theta(x)$.

Principle

First, $\bar{F}(y|x)$ is estimated by a kernel method. For all $(x, y) \in E \times \mathbb{R}$, let

$$\hat{\bar{F}}_n(y|x) = \sum_{i=1}^{n} K(d(x, X_i)/h_n)\mathbb{I}\{Y_i > y\} \Bigg/ \sum_{i=1}^{n} K(d(x, X_i)/h_n),$$

where

- h_n is a nonrandom sequence (called bandwidth) such that $h_n \to 0$ as $n \to \infty$,
- K is assumed to have a support included in $[0, 1]$ such that $C_1 \leq K(t) \leq C_2$ for all $t \in [0, 1]$ and $0 < C_1 < C_2 < \infty$.
  It is assumed without loss of generality that K integrates to one.
  K is called a type I kernel, see [3], Definition 4.1.

Second, $q(\alpha|x)$ is estimated via the generalized inverse of $\hat{\bar{F}}_n(.|x)$:

$$\hat{q}_n(\alpha|x) = \hat{\bar{F}}_n^{\leftarrow}(\alpha|x) = \inf\{y,\ \hat{\bar{F}}_n(y|x) \leq \alpha\},$$

for all $\alpha \in (0, 1)$.

Notations:

- $B(x, h_n)$ the ball of center x and radius h_n,
- $\varphi_x(h_n) := \mathbb{P}(X \in B(x, h_n))$ the small ball probability of X,
- $\mu_x^{(\tau)}(h_n) := \mathbb{E}\{K^\tau(d(x, X)/h_n)\}$ the τ-th moment,
- $\Lambda_n(x) = (n\alpha_n(\mu_x^{(1)}(h_n))^2/\mu_x^{(2)}(h_n))^{-1/2}$.

It is easily shown that for all $\tau > 0$

$$0 < C_1^\tau \varphi_x(h_n) \leq \mu_x^{(\tau)}(h_n) \leq C_2^\tau \varphi_x(h_n),$$

and thus $\Lambda_n(x)$ is of order $(n\alpha_n\varphi_x(h_n))^{-1/2}$.

Regularity assumption

Since the considered estimator involves a smoothing in the x direction, it is necessary to assess the regularity of the conditional survival function with respect to x. To this end, the oscillations are controlled by

$$\Delta\bar{F}(x,\alpha,\zeta,h) := \sup_{(x',\beta)\in B(x,h)\times[\alpha,\zeta]} \left| \frac{\bar{F}(q(\beta|x)|x')}{\bar{F}(q(\beta|x)|x)} - 1 \right|$$
$$= \sup_{(x',\beta)\in B(x,h)\times[\alpha,\zeta]} \left| \frac{\bar{F}(q(\beta|x)|x')}{\beta} - 1 \right|,$$

where $(\alpha,\zeta) \in (0,1)^2$.

Asymptotic normality

Theorem 1

Suppose **(A.1)**, **(A.2)** hold.

- Let $0 < \tau_J < \cdots < \tau_1 \leq 1$ where J is a positive integer.
- $x \in E$ such that $\varphi_x(h_n) > 0$ where $h_n \to 0$ as $n \to \infty$.

If $\alpha_n \to 0$ and there exists $\eta > 0$ such that $n\varphi_x(h_n)\alpha_n \to \infty$,

$$n\varphi_x(h_n)\alpha_n(\Delta\bar{F})^2(x, (1-\eta)\tau_J\alpha_n, (1+\eta)\alpha_n, h_n) \to 0,$$

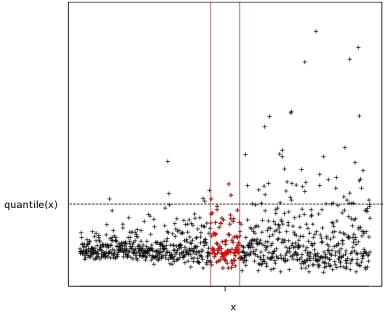
then, the random vector

$$\left\{\log(1/\alpha_n)\Lambda_n^{-1}(x)\left(\frac{\hat{q}_n(\tau_j\alpha_n|x)}{q(\tau_j\alpha_n|x)} - 1\right)\right\}_{j=1,\ldots,J}$$

is asymptotically Gaussian, centered, with covariance matrix $\theta^2(x)\Sigma$ where $\Sigma_{j,j'} = \tau_{j\wedge j'}^{-1}$ for $(j,j') \in \{1,\ldots,J\}^2$.

Conditions on the sequences

$n\varphi_x(h_n)\alpha_n \to \infty$: Necessary and sufficient condition for the almost sure presence of at least one point in the region $B(x, h_n) \times [q(\alpha_n|x), +\infty)$ of $E \times \mathbb{R}$.



$n\varphi_x(h_n)\alpha_n(\Delta\bar{F})^2(x, (1-\eta)\tau_J\alpha_n, (1+\eta)\alpha_n, h_n) \to 0$: The biais induced by the smoothing is negligible compared to the standard-deviation.

Principle

We propose a family of estimators of $\theta(x)$ based on some properties of the log-spacings of the conditional quantiles. Recall that

$$q(e^{-y}|x) = H^{-1}(y|x) = y^{\theta(x)}\ell(y|x).$$

Let $\alpha \in (0,1)$ small enough and $\tau \in (0,1)$,

$$\begin{aligned}
& \log q(\tau\alpha|x) - \log q(\alpha|x) \\
= \quad & \log H^{-1}(-\log(\tau\alpha)|x) - \log H^{-1}(-\log(\alpha)|x) \\
= \quad & \theta(x)(\log_{-2}(\tau\alpha) - \log_{-2}(\alpha)) + \log\left(\frac{\ell(-\log(\tau\alpha)|x)}{\ell(-\log(\alpha)|x)}\right) \\
\approx \quad & \theta(x)(\log_{-2}(\tau\alpha) - \log_{-2}(\alpha)) \\
\approx \quad & \theta(x)\frac{\log(1/\tau)}{\log(1/\alpha)},
\end{aligned}$$

where $\log_{-2}(\cdot) := \log\log(1/\cdot)$, Hence, for a decreasing sequence $0 < \tau_J < \cdots < \tau_1 \leq 1$, where J is a positive integer, and for all (twice differentiable) function $\phi : \mathbb{R}^J \to \mathbb{R}$ satisfying the shift and location invariance condition

$$\begin{aligned} \phi(\eta z) &= \eta\phi(z), \\ \phi(\eta u + z) &= \phi(z), \end{aligned}$$

for all $\eta \in \mathbb{R} \setminus \{0\}$, $z \in \mathbb{R}^J$ and where $u = (1, \ldots, 1)^t \in \mathbb{R}^J$, one has:

$$\theta(x) \approx \log(1/\alpha) \frac{\phi(\log q(\tau_1\alpha|x), \ldots, \log q(\tau_J\alpha|x))}{\phi(\log(1/\tau_1), \ldots, \log(1/\tau_J))}.$$

Thus, the estimation of $\theta(x)$ relies on the estimation of conditional quantiles $q(\cdot|x)$:

$$\hat{\theta}_n(x) = \log(1/\alpha_n) \frac{\phi(\log \hat{q}_n(\tau_1\alpha_n|x), \ldots, \log \hat{q}_n(\tau_J\alpha_n|x))}{\phi(\log(1/\tau_1), \ldots, \log(1/\tau_J))},$$

with $\alpha_n \to 0$ as $n \to \infty$.

Asymptotic normality

Theorem 2

Ssuppose **(A.1)–(A.3)** hold. Let $x \in E$ such that $\varphi_x(h_n) > 0$ where $h_n \to 0$ as $n \to \infty$. If $\alpha_n \to 0$,

$$\sqrt{n\varphi_x(h_n)\alpha_n}\varepsilon(\log(1/\alpha_n)|x) \to \lambda \in \mathbb{R}$$

and there exists $\eta > 0$ such that $n\varphi_x(h_n)\alpha_n \to \infty$ and

$$\sqrt{n\varphi_x(h_n)\alpha_n}\{\Delta\bar{F}(x,(1-\eta)\tau_J\alpha_n,(1+\eta)\alpha_n,h_n) \vee 1/\log(1/\alpha_n)\} \to 0,$$

then,

$$\Lambda_n^{-1}(x)(\hat{\theta}_n(x) - \theta(x)) \xrightarrow{d} \mathcal{N}(\mu_\phi, \theta^2(x)V_\phi)$$

where $\mu_\phi = \lambda v^t \, \nabla \log \phi(v)$, $V_\phi = (\nabla \log \phi(v))^t \, \Sigma \, (\nabla \log \phi(v))$ and $v = (\log(1/\tau_1), \ldots, \log(1/\tau_J))^t$ do not depend on (X, Y) distribution.

Corollary

Suppose **(A.1)–(A.3)** hold. Let $x \in E$ such that $\varphi_x(h_n) > 0$ and $y\varepsilon(y|x) \to \infty$ as $y \to \infty$. Assume there exist L_θ, L_c et L_ε such that

$$\begin{aligned} \left|\frac{1}{\theta(x)} - \frac{1}{\theta(x')}\right| &\leq L_\theta d(x,x'), \\ |\log c(x) - \log c(x')| &\leq L_c d(x,x'), \\ \sup_{u \in [1, \bar{y}_n(x)]} |\varepsilon(u|x) - \varepsilon(u|x')| &\leq L_\varepsilon d(x,x'), \end{aligned}$$

where $\bar{y}_n(x) := \sup\{H(q(\alpha_n|x)|x'),\ x' \in B(x,h_n)\}$. Suppose

$$\varphi_x^{-1}(1/y)(\log y)^{1+\xi-\rho(x)} \to 0 \tag{2}$$

for some $\xi > 0$ as $y \to \infty$. Then, letting $\lambda > 0$,

$$\alpha_n = n^{-1+\xi} \text{ and } h_n = \varphi_x^{-1}\left(\lambda(1-\xi)^{2\rho(x)} n^{-\xi} (\varepsilon(\log n|x))^{-2}\right),$$

Theorem 2 yields $\Lambda_n^{-1}(x)(\hat{\theta}_n(x) - \theta(x)) \xrightarrow{d} \mathcal{N}(\mu_\phi, \theta^2(x)V_\phi)$.

The key assumption (2) holds in the finite dimensional setting or for fractal-type and some exponential-type processes, see [3], Chapter 13.

Let us focus on the functions $\phi^{(p)}(z) = \left(\sum_{j=2}^{J} \beta_j (z_j - z_1)^p\right)^{1/p}$, where $z = (z_1, \ldots, z_J)^t \in \mathbb{R}^J$, $p \in \mathbb{N} \setminus \{0\}$ and for all $j \in \{2, \ldots, J\}$, $\beta_j \in \mathbb{R}$.
The corresponding estimator of θ writes:

$$\hat{\theta}_n^{(p)}(x) = \log(1/\alpha_n) \left(\frac{\sum_{j=2}^{J} \beta_j \left[\log \hat{q}_n(\tau_j \alpha_n | x) - \log \hat{q}_n(\tau_1 \alpha_n | x)\right]^p}{\sum_{j=2}^{J} \beta_j [\log(\tau_1/\tau_j)]^p} \right)^{1/p}.$$

As a consequence of Theorem 2, the associated asymptotic mean and variance of $\hat{\theta}_n^{(p)}(x)$ are given for an arbitrary vector β by $\mu = \lambda$ and

$$V^{(p)} = \frac{(\eta^{(p)})^t A \Sigma A^t \eta^{(p)}}{(\eta^{(p)})^t A v v^t A^t \eta^{(p)}},$$

where A is a given matrix and $\eta^{(p)} = (\beta_j(v_j - v_1),\ j = 2, \ldots, J)^t$.

- The asymptotic bias μ does not depend neither on p and nor on the weights $\{\beta_j,\ j = 2, \ldots, J\}$.
- It is possible to minimize $V^{(p)}$ with respect to $\eta^{(p)}$.

Proposition

The asymptotic variance of $\hat{\theta}_n^{(p)}(x)$ is minimal for $\eta^{(p)}$ proportional to $\eta_{\rm opt} = (A\Sigma A^t)^{-1}Av$ and is given by

$$V_{\rm opt} = \frac{1}{(Av)^t\ (A\Sigma A^t)^{-1}Av},$$

and is independent of p.

Moreover, for a fixed value of J, it is possible to minimize numerically the optimal variance $V_{\rm opt}$ with respect to parameters $0 < \tau_J < \cdots < \tau_1 \leq 1$. The resulting values of $V_{\rm opt}$ are displayed in the table below:

J	$V_{\rm opt}$	τ_1	τ_2	τ_3	τ_4	τ_5
2	1.5441	1.0000	0.2032			
3	1.2191	1.0000	0.3615	0.0735		
4	1.1223	1.0000	0.4703	0.1702	0.0346	
5	1.0789	1.0000	0.5486	0.2585	0.0936	0.0190

- E is a subset of $L^2([0,1])$ made of trigonometric functions $\psi_z : [0,1] \to [0,1]$, $\psi_z(t) = \cos(2\pi z t)$ with different periods indexed by $z \in [1/10, 1/2]$.

- Two semi-metrics are considered:

$$\begin{aligned} d_1(\psi_z, \psi_{z'}) &= \left| \|\psi_z\|_2^2 - \|\psi_{z'}\|_2^2 \right|, \\ d_2(\psi_z, \psi_{z'}) &= \|\psi_z - \psi_{z'}\|_2, \end{aligned}$$

for all $(z, z') \in [1/10, 1/2]^2$, where

$$\|\psi_z\|_2^2 = \int_0^1 \psi_z^2(t)dt = \frac{1}{2}\left(1 + \frac{\sin(4\pi z)}{4\pi z}\right).$$

The semi-metric d_2 is built on the classical L_2 norm while d_1 measures some spacing between the periods of the trigonometric functions.

Simulated data

$N = 100$ copies of a sample of size $n = 1000$ from a random pair (X, Y) defined as follows:

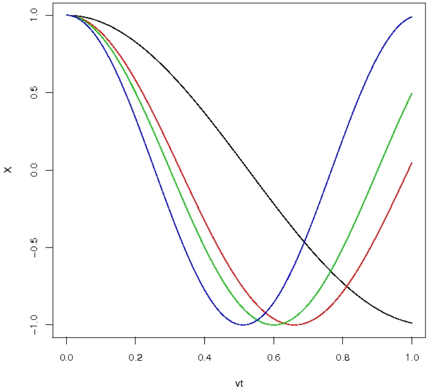
- The covariate X is chosen randomly on E by considering $X = \psi_Z$ where Z is a uniform random variable on $[1/10, 1/2]$.
- For a fixed function $x \in E$, the generalized inverse of the conditional hazard function $H(.|x)$ is given by the following Hall's model:

$$H^{\leftarrow}(y|x) = y^{\theta(x)} \left(1 - \gamma y^{\rho(x)}\right), \; y \geq 0,$$

with

$$\begin{aligned} \theta(x) &= (18/5\|x\|_2^2 + 9/50)^{-1} - 5/18, \\ \rho(x) &= 50/(60\|x\|_2^2 + 3) - 5/2, \\ \gamma &= 1/10. \end{aligned}$$

Four simulated random functions $X(.)$



Estimators

- The previous estimator with optimal weights is used. Here, we limit ourselves to $J = 5$ and $p \in \{1, 3\}$.
- A modified bi-quadratic kernel is adopted (type I kernel):

$$K(u) = \frac{10}{9}\left(\frac{3}{2}\left(1 - u^2\right)^2 + \frac{1}{10}\right)\mathbb{I}\{|u| \leq 1\}.$$

- h_n and α_n are selected simultaneously thanks to a data-driven procedure. For a fixed x, let $\{Z_1(x, h_n), \ldots, Z_{m_n}(x, h_n)\}$ be the m_n random values Y_i for which $X_i \in B(x, h_n)$. The idea [7] is to select the sequences h_n and α_n such that the rescaled log-spacings

$$i \log(m_n/i)(\log Z_{m_n-i+1,m_n}(x, h_n) - \log Z_{m_n-i,m_n}(x, h_n)),$$

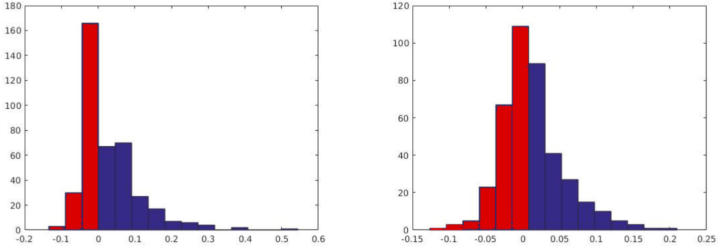
$i = 1, \ldots, \lfloor m_n\alpha_n \rfloor$, are approximately $Exp(\theta(x))$ distributed. The "optimal" sequences are obtained by minimizing a Kolmogorov-Smirnov distance.

- Comparison with the non-conditional estimator proposed in [2]:

$$\hat{\theta}_n^{NCE} = \frac{\sum_{i=1}^{k_n}\left(\log Y_{n-i+1,n} - \log Y_{n-k_n+1,n}\right)}{\sum_{i=1}^{k_n}\left(\log_{-2}(n/i) - \log_{-2}(n/k_n)\right)}.$$

Influence of the exponent p

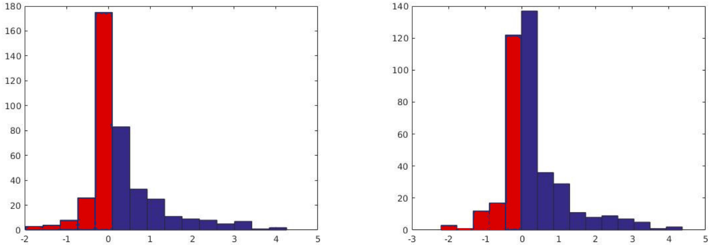
Let $e_{i,\ell,p}$ be the relative error obtained on the ith replication using the semi-metric d_ℓ and the estimator $\hat{\theta}^{(p)}$.



Left: histogram of $e_{\bullet,1,3} - e_{\bullet,1,1}$ (semi-metric d_1), right: histogram of $e_{\bullet,2,3} - e_{\bullet,2,1}$ (semi-metric d_2).
Both histograms are nearly centered, small influence of p.

Influence of the semi-metric

Recall that $e_{i,\ell,p}$ is the relative error obtained on the ith replication using the semi-metric d_ℓ and the estimator $\hat{\theta}^{(p)}$.

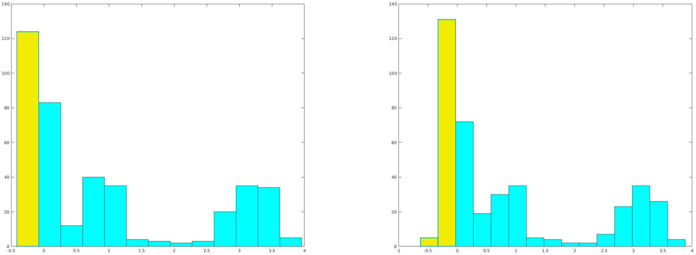


Left: histogram of $e_{\bullet,2,1} - e_{\bullet,1,1}$ ($p = 1$), right: histogram of $e_{\bullet,2,3} - e_{\bullet,1,3}$ ($p = 3$).

Both histograms are skewed to the right, the semi-metric d_1 yields better result than d_2.

Comparison with the non-conditional estimator

Recall that $e_{i,\ell,p}$ is the relative error obtained on the ith replication using the semi-metric d_ℓ and the estimator $\hat{\theta}^{(p)}$. We moreover denote by e_i the relative error obtained on the ith replication using the non-conditional estimator $\hat{\theta}_n^{NCE}$.



Left: histogram of $e_\bullet - e_{\bullet,1,1}$ ($p = 1$), right: histogram of $e_\bullet - e_{\bullet,1,3}$ ($p = 3$).
Both histograms are skewed to the right, the conditional estimator yields better results than the unconditional one.

References

1. Bingham, N.H., Goldie, C.M., Teugels, J.L. (1987) *Regular Variation*, Cambridge University Press.
2. Daouia, A., Gardes, L., Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19, 2557–2589.
3. Ferraty, F., Vieu, P. (2006). *Nonparametric functional data analysis*, Springer.
4. Gardes, L., Girard, S. (2006). Comparison of Weibull tail-coefficient estimators. *REVSTAT - Statistical Journal*, 4(2), 163–188.
5. Gardes, L., Girard, S. (2012). Functional kernel estimators of large conditional quantiles. *Electronic Journal of Statistics*, 6, 1715–1744.
6. Gardes, L., Girard, S. (2016). On the estimation of the functional Weibull tail-coefficient. *Journal of Multivariate Analysis*, 146, 24–45.
7. Gardes, L., Girard, S., Lekina, A. (2010). Functional nonparametric estimation of conditional extreme quantiles. *Journal of Multivariate Analysis*, 101, 419–433.